From the
DEPARTMENT OF MICROBIOLOGY, TUMOR AND CELL BIOLOGY
Karolinska Institutet, Stockholm, Sweden

ANTIGENIC VARIATION IN *PLASMODIUM FALCIPARUM*: UNDERSTANDING THE RIFIN PROTEIN FAMILY

Nicolas Joannin

Karolinska Institutet

Stockholm 2010

Published by Karolinska University Press; Printed by Larserics Digital Print AB.
Box 200, SE-171 77 Stockholm, Sweden

ISBN 978-91-7457-075-05

ABSTRACT

RIFIN proteins are variable surface antigens, which have a central role in the survival and virulence of the malaria parasite *Plasmodium falciparum*. Antigenic variation is a mean for these parasites to avoid clearance by the host's immune system. However, this is often a secondary function to the main role of these proteins. In the case of RIFIN, *P. falciparum*'s largest multicopy protein family, the main functions remain unknown. In order to elucidate a protein's function, it is crucial to understand its basic properties, including the structure of the protein family, their localization and the protein's topology.

Through different methods, we have strived to simplify the RIFIN protein family into manageable entities. We have started with a simple classification of RIFIN proteins into meaningful sub-groups. We have predicted that these sub-groups are functionally distinct, although they probably share a related function.

We then designed RSPred, an automatic method, based on hidden Markov models and a sorting program, to detect and classify RIFIN and STEVOR sequences according to their sub-group.

Finally, using an *in vitro* method to determine protein topology, we have analyzed both A-RIFIN and B-RIFIN proteins for their number of transmembrane segments and their topologies. We show that both protein groups have a signal sequence targeting them to lipid bilayers and only one transmembrane domain. They both share a common topology where the bulk of the protein is exposed to the extracellular environment.

With the current knowledge of RIFIN protein localizations, as well as the loss of expression of A-RIFIN but not B-RIFIN proteins in a splenectomized patient, it seems increasingly clear that B-RIFIN proteins are good targets for future studies, to decipher the functions of these variable proteins.

LIST OF PUBLICATIONS

I. Joannin N, Abhiman S, Sonnhammer E, Wahlgren M: **Sub-grouping and sub-functionalization of the RIFIN multi-copy protein family**. *BMC Genomics* 2008, **9**(1):19.

II. Joannin N*, Kallberg Y*, Wahlgren M, Persson B: **RSpred, a set of hidden Markov models to detect and classify the RIFIN and STEVOR proteins of *Plasmodium falciparum***. *Submitted manuscript*. Shared first authorship.

III. Joannin N, Enquist K, Wahlgren M, von Heijne G, Nilsson I: **RIFIN topology: a new perspective on *Plasmodium falciparum* variable surface antigens**. *Manuscript*.

IV. Hayes C, Diez D, Joannin N, Honda W, Kanehisa M, Wahlgren M, Wheelock C, Goto S: **varDB: a pathogen-specific sequence database of protein families involved in antigenic variation**. *Bioinformatics* 2008, **24**(21):2564-5.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

WHO	World Health Organization
GMEP	Global Malaria Eradication Program
RBC	Red Blood Cell
PfEMP1	*Plasmodium falciparum* Erythrocyte Membrane Protein 1
Pfmc-2TM	*P. falciparum* Maurer's cleft-2 transmembrane
Stevor	Subtelomeric variable open reading frame
Rif	Repetitive interspersed family
TM	Transmembrane
DBL	Duffy Binding Like
CSS	Conservation Shifting Site
RSS	Rate Shifting Site

1 INTRODUCTION

My thesis is mostly about a family of antigenic variant proteins called RIFIN, which are uniquely found in the parasite that causes the most dangerous form of malaria, *Plasmodium falciparum*. But before I go into the details of my work on this protein family, I will briefly introduce malaria in general, the *Plasmodium* parasites, as well as antigenic variation and specific protein families.

1.1 MALARIA

Among the top infectious diseases in the world malaria comes in third place, after tuberculosis and HIV, in terms of annual casualties. Today, this disease extends from Afghanistan to South Africa, including many Asian and South American countries. There is an estimated 3 billion people at risk of this disease and 250 million cases that result in approximately 1 million deaths every year (WHO 2008). Although these estimates cover 109 countries and territories, the bulk of the burden is attributed to Africa (Figure 1, (Hay, Guerra et al. 2009)), where the majority of deaths concern children under the age of 5 and pregnant women.

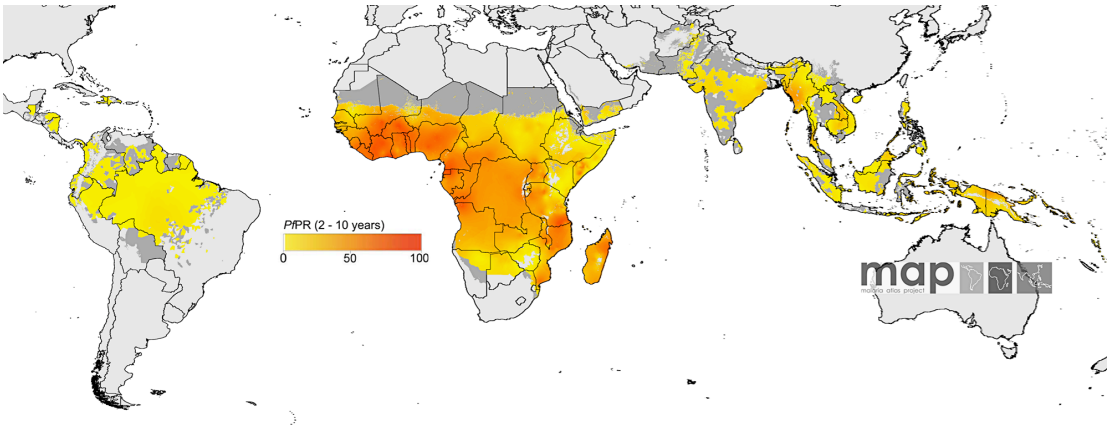

Figure 1: Stable spacial limits of *P. falciparum* malaria transmission (low to high, from yellow to red), low risk of transmission in dark grey, no risk in light grey. From (Hay, Guerra et al. 2009).

In the mid-twentieth century, malaria had a broader range than today. In 1955, the World Health Organization launched the Global Malaria Eradication Program (GMEP). This campaign was successful in eliminating malaria from Europe, North Africa, North America, Australia and parts of Asia, the Middle East and South America. However, the disease persisted in the most

economically poor regions of the world (Sachs and Malaney 2002). The eradication program was abandoned and replaced with a long-term focus on control of the disease in 1969 (Snow and Marsh 2010).

The reasons for this failure to eliminate malaria in the 1950's and 1960's is ascribed to many factors, including ecological, socio-political, economical and biological. They are manifold; here are a few (simplified) examples: First, the regions that were successful in eliminating malaria had temperate climates with seasonal transmission, which allowed for concentrated efforts in the end of the malaria season. Second, the political unrest in many countries in sub-Saharan Africa, as they became independent, gave little support to the GMEP. Third, the very effective drug chloroquine was not used to treat native populations of endemic areas because it was deemed too expensive to distribute. Finally, the development of resistance to drugs and insecticides as well as the absence of any effective vaccine resulted in a failure of the GMEP.

The two following decades of relaxed efforts to control malaria and increased drug resistance have resulted in a marked resurgence of disease and deaths due to malaria (Sachs 2002). Since the mid-1990's, several initiatives, networks and foundations have boosted the efforts in malaria control and research. These include, non-exhaustively, the Multilateral Initiative on Malaria (MIM), the African Malaria Vaccine Testing Network (AMVTN), ACTMalaria-Asia, Roll Back Malaria (RBM), Medicines for Malaria Venture (MMV), Malaria Vaccine Initiative (MVI) and The Bill & Melinda Gates Foundation. Additionally, scientific consortiums have allowed for the sequencing of several *Plasmodium* and *Anopheles* genomes and the creation of the Malaria Research and Reference Reagent Resource Center (MR4).

These renewed efforts and the advances in malaria research have brought back the eradication of malaria to the public debate (Tanner and de Savigny 2008; Cohen, Moonen et al. 2010). Still, much effort will be needed to accomplish this elusive feat. As parasites develop novel resistance mechanisms to existing drug therapies, current research continues to discover new drugs to compete in this arms race (Dharia, Plouffe et al. 2010; Rottmann, McNamara et al. 2010). Additionally, although there is hope for the licensing of a vaccine that would be effective against clinical and severe disease (Casares, Brumeanu et al. 2010), sterile immunity remains elusive,

which might allow for the parasite to evolve a counter attack. Unfortunately, this arms race may be an interminable one that will be financially unsustainable (Sinden 2010).

1.2 PLASMODIUM SPECIES

There are over 200 *Plasmodium* species, which all infect vertebrate animals; however, only five infect humans. These are *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and *P. knowlesi*. They have a complex life cycle that involves mosquitoes and primates, with several distinct morphological and physiological stages. The mosquitoes are of the genus *Anopheles* and the primates are primarily *Homo sapiens*, although other primates have been found to be susceptible to infections (Krief, Escalante et al. 2010). One exception is *Plasmodium knowlesi*, which was initially thought to infect only macaque monkeys, but has recently been shown to be widely distributed and life threatening to humans (Cox-Singh, Davis et al. 2008; Cox-Singh, Hiu et al. 2010).

These parasites are protozoa of the *Apicomplexan* phylum. They are haploid organisms during most of their life cycle, only becoming diploid briefly in the gut of the mosquito before undergoing meiosis. Therefore, within each host, the propagation of the parasite is accomplished through asexual multiplication. Figure 2 depicts the life cycle of *Plasmodium falciparum*, which is roughly similar to that of the other species.

Briefly, from the perspective of humans, the cycle “starts” when an infected mosquito bites a human and injects, from its salivary glands, sporozoites into the skin. These sporozoites then migrate to blood capillaries in order to reach the liver, where they invade hepatocytes. The parasites differentiate and undergo a first asexual multiplication stage during approximately 14 days, resulting in tens of thousands of merozoites that burst from the infected hepatocytes.

Released into the blood flow, these merozoites initiate the erythrocytic cycle of asexual multiplication. This cycle starts with the invasion of red blood cells (RBC) and lasts approximately 48 hours. Once the parasite has penetrated the RBC, it matures from ring stage trophozoite to mature trophozoite, during which it grows and prepares for chromosome replication. The transition from

trophozoite to schizont is marked by the first mitosis and schizogony continues until the last mitosis. Merogony is the assembly of nuclei and organelles into the merozoites. This late stage can then burst, releasing 8 to 26 new merozoites. Parasite numbers increase exponentially in the blood stream of infected individuals until either the host's immune system or chemotherapy controls its progress, or until the human host dies.

Some parasites may take an alternative developmental pathway, transitioning from trophozoites into gametocytes (instead of schizonts), male and female sexual forms, that may then be taken up by a feeding mosquito.

Within the mosquito gut, female and male gametocytes transform into one macro- and eight microgametes, respectively. One microgamete and one macrogamete can then fuse, mature and complete meiosis, to form a motile ookinetes that migrate through the cells of the stomach wall. Once this barrier crossed it encysts, creating a wart-like structure known as the oocyst. For the third iteration in the parasite's life cycle, it proceeds to asexual multiplication resulting in the liberation of thousands of sporozoites that migrate to the mosquito's salivary glands. Thus, the *Plasmodium falciparum* life cycle is completed... and starts over again, when the mosquito bites another human.

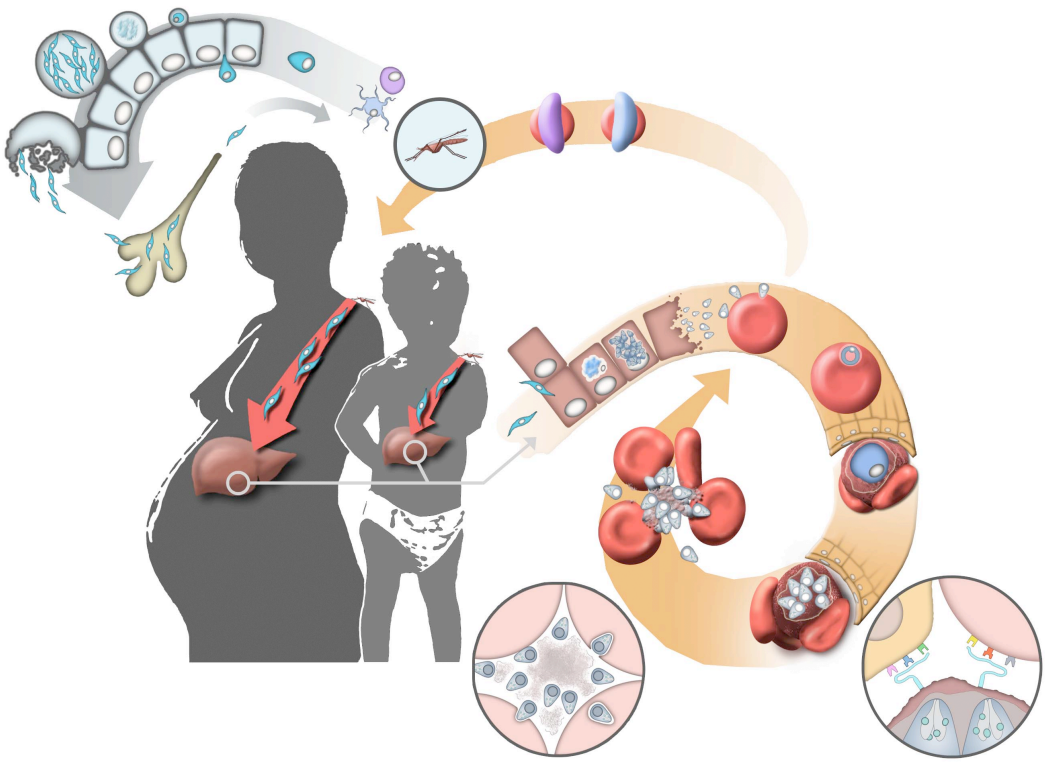

Figure 2: Life cycle of the *Plasmodium falciparum* parasite. Courtesy of Dr. Johan Normark.

1.3 *PLASMODIUM FALCIPARUM*'S SURVIVAL WITHIN THE HUMAN HOST: PFEMP1, AN ANTIGENIC VARIANT PROTEIN FAMILY

As a species, a successful infectious agent is an organism that manages to survive long enough within its host to spread to other hosts. However, vertebrates have developed an elaborate immune system that efficiently targets the non-self. In order to extend the length of infection, pathogens have evolved mechanisms to avoid being destroyed by their host's immune system. Some, like *Toxoplasma*, invade and highjack host cells in order to hide from opsonizing antibodies and destruction, as well as to disseminate within the host (Lambert and Barragan 2010). However, *P. falciparum* has evolved its life cycle so as to inhabit a terminally differentiated de-nucleated human cell: the red blood cell. Contrary to most other cells in the human body, RBCs are devoid of almost all the features that allows *Toxoplasma* to survive within other cells. Therefore, the malaria parasite has to modify the RBC in order to feed and eliminate waste, as well as to export proteins for the evasion of host defenses (reviewed in (Goldberg and Cowman 2010)). Also, it is continuously multiplying, destroying and invading new RBCs every 48 hours, necessarily exposing an array of proteins involved in these processes (Ferreira, Zilversmit et al. 2007). These modifications to the RBC make it vulnerable to the adaptive immunity and to clearance by the spleen, which removes damaged and old RBCs from the circulation (David, Hommel et al. 1983; Udeinya, Miller et al. 1983).

In such an environment, this parasite has to cope with two strong selective pressures: First, clearance by the spleen. Second, the generation of antibodies specific to the antigens it exposes at the surface of sporozites, merozoites and infected RBCs. One mechanism that addresses both pressures, in the infected RBCs, is antigenic variation of the *Plasmodium falciparum* Erythrocyte Membrane Protein 1 (PfEMP1) (Baruch, Pasloske et al. 1995; Smith, Chitnis et al. 1995; Su, Heatwole et al. 1995).

The PfEMP1 proteins have been identified as the main ligands responsible for cytoadherence (Sherman, Eda et al. 2003) and rosetting (Rowe, Moulds et al. 1997; Chen, Barragan et al. 1998). These phenomena result in the sequestration of late trophozoite and schizont infected RBCs in the deep

vasculature of host tissues, enabling the parasite to avoid clearance by the spleen (MacPherson, Warrell et al. 1985; Pongponratn, Riganti et al. 1985; Turner, Morrison et al. 1994; Silamut, Phu et al. 1999).
However, expression of the PfEMP1 protein for such extended periods of time makes it the target of naturally acquired immunity (Bull, Lowe et al. 1998; Bull, Lowe et al. 1999). To counter the host's ability to elaborate custom defenses against specific protein variants, the parasite has expanded the PfEMP1 encoding genes, named *var* genes, to a polymorphic family of ~60 members per haploid genome (84 in 3D7 when pseudogenes are included). It switches the expression from one member of the family to another, in a mutually exclusive manner, only exposing one protein at the infected RBC surface at a time (Chen, Fernandez et al. 1998; Scherf, Hernandez-Rivas et al. 1998).

This dogma has recently been questioned (Brolin, Ribacke et al. 2009; Sander, Salanti et al. 2009; Joergensen, Bengtsson et al. 2010); however, the experiments that show co-expression of two PfEMP1 proteins at the surface of the infected RBC are the result of *in vitro* selection. This situation remains to be shown for fresh isolates from patients (Joergensen, Bengtsson et al. 2010). Also, it is unlikely that a large number of PfEMP1 proteins would be exposed simultaneously at the surface of the infected RBC as that would result in a rapid exhaustion of the antigenic repertoire.

Although the sequestration of the infected RBC is posited to have evolved, in *P. falciparum,* to prevent its passage through the spleen (Deitsch, Lukehart et al. 2009), there is little direct evidence to support this. A recent case study of a splenectomized patient with *falciparum* malaria supports this idea (Bachmann, Esser et al. 2009). Bachmann *et al.* showed that the parasites found in the patient were unable to cytoadhere to various endothelial receptors and did not transcribe *var* genes.

Whether immune evasion is the primary role or a necessary side effect of PfEMP1 proteins, it appears clear that it has a central impact on the establishment of chronic infections and the survival of the species. Figure 3 shows a simplified diagram of the relationship between antigenic variation and pathogen transmission

1.4 ANTIGENIC VARIATION

Antigenic variation is often broadly defined as the capacity for a pathogen to change its surface proteins in order to evade a host immune response. Many infectious organisms fulfill this definition, including viruses, bacteria and protozoan parasites. However, these consist of different mechanisms such as phase variation, antigenic drift and shift, as well as antigenic variation *sensu stricto* (Mumford 2007; Deitsch, Lukehart et al. 2009).

Figure 3 represents a general idea of the link between survival, transmission and antigenic variation / antigenic switching (Turner 2002).

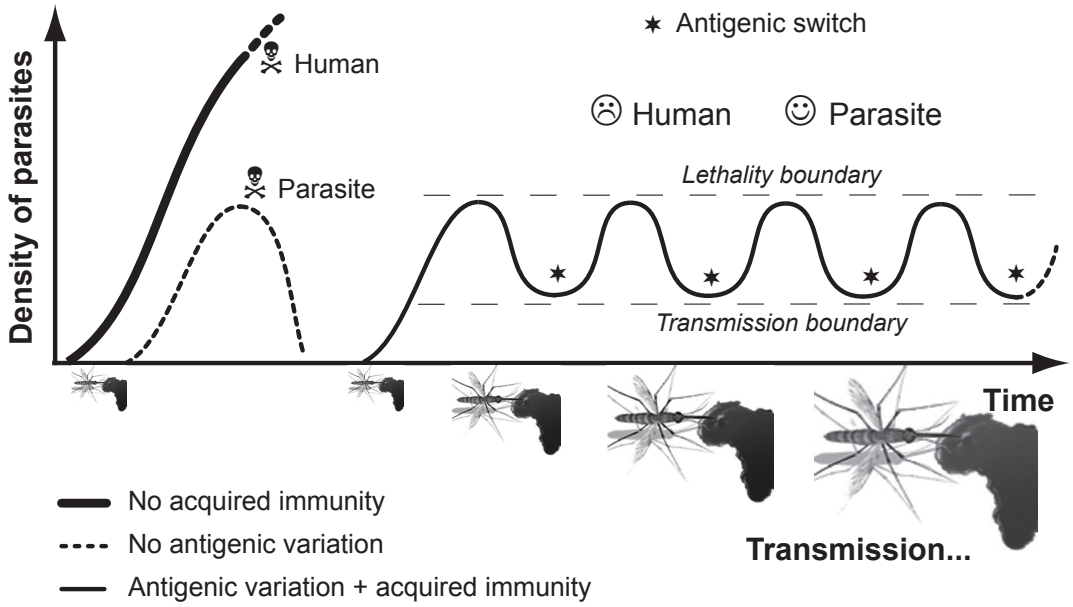

Figure 3: Simplified diagram of the relationship between antigenic variation and pathogen transmission (Adapted from Turner, 2002)

Phase variation is mostly an attribute of some bacteria that undergo frequent and usually reversible phenotypic changes resulting from genetic or epigenetic alterations at specific loci. It is characteristically the switch ON / switch OFF of a specific phenotype (van der Woude and Bäumler 2004; Wisniewski-Dyé and Vial 2008). This is used by bacteria to generate intra-population diversity that is important to escape host defenses or in niche adaptation.

Antigenic drift and shift are usually applied to viruses. Antigenic drift is the rapid accumulation of mutations during replication. The term antigenic shift is sometimes used to express a leap in the modification of the surface protein due to recombination events (Mumford 2007). In order to avoid clearance by hosts' immune system or simply to allow for re-infection, the virus needs to vary its antigenic proteins. These mutated forms are often referred as immune-escape variants because they have accumulated a sufficient number of amino acid changes to avoid recognition by vaccinated and unvaccinated hosts alike (Boni 2008).

Antigenic variation *sensu stricto* is usually characterized as the switching between members of a multicopy gene family, which involves a complicated mechanism of activation and silencing (Craig and Scherf 2003). C.M. Turner suggests the following requirements to fulfill the definition of antigenic variation (Turner 2002):

- Capacity to interact with the environment;
- Mutually exclusive expression of variable antigens in each cell within an infection;
- Mutually exclusive expression in the within-host pathogen population;
- Protein variants must be antigens confronted with functional immunity;
- Capacity for population growth within a host.

Interestingly, these definitions do not constrain the function of antigenic variant proteins. Indeed, some parasites, such as the *Trypanosoma*, have developed antigenic variation to fulfill the main function of immune evasion (Morrison, Marcello et al. 2009). However, as we have seen in *Plasmodium falciparum*, the PfEMP1 protein family couples the immune evasion function with a binding function, which allows the parasite to sequester and avoid splenic clearance.

Unfortunately, the requirements of Turner's strict definition may not always be practical to evaluate, making it difficult to determine whether a multicopy protein family is antigenic variant or not. Also, the first condition would exclude the free-living organisms that express antigenic variants (Simon and

Schmidt 2007) as they would not induce “immunity”. Nevertheless, they suggest the possibility of the recognition of specific surface antigens by the ciliate predator *Didinium nasutum*, which would explain the need for antigenic variation in the absence of immunity (Simon and Schmidt 2007).

1.5 ANTIGENIC VARIATION AND THE OTHER VARIANT PROTEIN FAMILIES OF *PLASMODIUM FALCIPARUM*

In *P. falciparum,* antigenic variation is often solely ascribed to the PfEMP1 protein family, which is described as the major surface molecule of the infected RBC (Barry, Leliwa-Sytek et al. 2007; Frank and Enderes 2010). As a result, much effort has been focused on the PfEMP1 protein family (a PubMed search for PfEMP1 yields 272 versus 74 for all three of the other multicopy families). However, *P. falciparum* exposes three other multicopy variant protein families at the surface of the infected RBC: PfMC-2TM, STEVOR and RIFIN. These proteins are encoded by the *pfmc-2tm*, *stevor* and *rif* genes, short for *Plasmodium falciparum Maurer’s Cleft two transmembrane* protein, *subtelomeric variable open reading frame* and *repetitive interspersed family,* respectively (Cheng, Lawrence et al. 1997; Cheng, Cloonan et al. 1998; Fernandez, Hommel et al. 1999; Kyes, Rowe et al. 1999; Sam-Yellowe, Florens et al. 2004).

According to the most current gene annotations found at www.genedb.org, there are 12 protein coding *pfmc-2tm* genes. The *stevor* are a small family of 31 protein coding genes and *rif* is the largest gene family of *P. falciparum* with 161 protein coding genes.

All three protein families have been described as variable surface antigens (Fernandez, Hommel et al. 1999; Kyes, Rowe et al. 1999; Lavazec, Sanyal et al. 2006; Niang, Yan Yam et al. 2009). In particular, RIFIN and STEVOR proteins are recognized by human antibodies that are induced during natural *P. falciparum* infections (Abdel-Latif, Khattab et al. 2002; Abdel-Latif, Dietz et al. 2003; Schreiber, Brattig et al. 2006; Schreiber, Khattab et al. 2008). However, this has not been unambiguously shown for PfMC-2TM proteins (Sam-Yellowe, Florens et al. 2004; Tsarukyanova, Drazba et al. 2009). It is therefore impossible to confirm interaction with the environment for this protein family.

Strict mutual exclusion of the expression of these proteins has not been proven, but there is evidence for the expression of a restricted number of variants (Lavazec, Sanyal et al. 2007; Wang, Magistrado et al. 2008; Cabral and Wunderlich 2009; Niang, Yan Yam et al. 2009). In the case of RIFIN and STEVOR proteins, it is clear that several variants can be expressed at the same time, in the same cell (Petter, Haeggström et al. 2007; Niang, Yan Yam et al. 2009). These proteins apparently do not fulfill the mutually exclusive protein expression criterium.

Finally, a recent publication has reported the absence of multicopy gene family expression in *P. falciparum* from a splenectomized malaria patient. More precisely they report on the apparent down regulation of *var*, *stevor* and a sub-type of *rif* genes (Bachmann, Esser et al. 2009). One of the authors' conclusions is that the expression of these proteins is required for sequestration and avoidance of splenic clearance. Therefore, in a normal context, the expression of these proteins would enhance the survival and growth of the parasite population. The authors have also reported on the maintenance of the expression of the *pfmc-2tm* genes and the other sub-type of *rif* genes. However, this does not necessarily imply that the expression of these proteins does not participate in the growth of the population. On the contrary, the maintenance of their expression marks their importance for the survival of the parasite. Hence, fulfilling Turner's last requirement.

As I mentioned above, Turner's requirements for a protein family to be antigenic variant, *sensu stricto*, are difficult to prove/disprove. However, if we adopt a broader definition of antigenic variation, it becomes reasonable to assign RIFIN and STEVOR sequences as antigenic variant protein families.

1.6 RIFIN AND STEVOR PROTEINS

The RIFIN protein family was first described as "interspersed repetitive DNA elements" (Weber 1988) and formally annotated as RIFIN in the publication of the chromosome 2 of *Plasmodium falciparum*, although, this name was used indiscriminately for both RIFIN and STEVOR proteins (Gardner, Tettelin et al. 1998). The STEVOR proteins were first reported as 7H8/6, multicopy genes found in all *P. falciparum* strains tested (Limpaiboon, Taylor et al. 1990).

In addition to the brief description of RIFIN and STEVOR proteins provided by Gardner (Gardner, Tettelin et al. 1998), Cheng (Cheng, Cloonan et al. 1998) also described the *stevor* and *rif* gene organization and deduced protein characteristics, suggesting that these sequences are part of a protein super family. According to their analyses, both *rif* and *stevor* genes are composed of two exons and have a subtelomeric location, centromeric to the *var* genes. The first exon is short and encodes a putative signal peptide or anchor, whereas the second exon encodes the bulk of the 27 to 35 kDa protein. Predictive methods estimated that these proteins possessed three transmembrane (TM) domains (Figure 3), the first of which corresponding to the predicted signal sequence.

Multiple sequence alignments, of the deduced RIFIN and STEVOR proteins available at the time, showed that these sequences were fairly conserved except for the region between the two predicted TM domains, which was highly variable. The deduced topology of the protein placed the relatively conserved N-terminal third of the protein, as well as the very conserved C-terminus, on the cytoplasmic side of the membrane and the highly polymorphic loop on the extracellular side. These observations as well as the detection of a limited number of transcripts, in the trophozoite and schizont stages of the erythrocytic cycle, led both Cheng *et al.* (Cheng, Cloonan et al. 1998) and Gardner *et al.* (Gardner, Tettelin et al. 1998) to suggest that *rif* and *stevor* genes may be clonally variant.

It was only a year later that RIFIN proteins were shown to be expressed at the surface of the infected RBC (Fernandez, Hommel et al. 1999; Kyes, Rowe et al. 1999). Eventually, the trafficking of RIFIN proteins through Maurer's cleft, as well as their localization in gametocytes and merozoites was also demonstrated (Haeggström, Kironde et al. 2004; Haeggström, VON Euler et al. 2007; Petter, Haeggström et al. 2007; Petter, Bonow et al. 2008). Conversely, the STEVOR proteins were initially found to localize to the Maurer's cleft compartment in the infected RBC cytoplasm, but not to the RBC surface (Kaviratne, Khan et al. 2002; McRobert, Preiser et al. 2004; Przyborski, Miller et al. 2005). However, García *et al.* (García, Puentes et al. 2005) described in 2005 the binding of STEVOR derived peptides to normal RBCs. Eventually, others found evidence of STEVOR proteins expressed at

the infected RBC and merozoite surfaces, as well as in gametocytes and sporozoites (McRobert, Preiser et al. 2004; Lavazec, Sanyal et al. 2006; Blythe, Yam et al. 2008; Khattab, Bonow et al. 2008; Niang, Yan Yam et al. 2009).

The development of large scale genomics, proteomics and microarray projects added to the evidence of transcription and protein expression of both RIFIN and STEVOR in multiple stages of the *P. falciparum* life cycle, including gametocytes, merozoites and sporozoites (Florens, Washburn et al. 2002; Le Roch, Zhou et al. 2002; Bozdech, Llinás et al. 2003; Le Roch, Zhou et al. 2003; Florens, Liu et al. 2004; Le Roch, Johnson et al. 2004; Daily, Le Roch et al. 2005; Llinás, Bozdech et al. 2006).

The publication of the genome sequence of the *Plasmodium falciparum* parasite allowed for genome wide analysis of the multicopy gene families (Gardner, Hall et al. 2002; Lavstsen, Salanti et al. 2003; Joannin, Abhiman et al. 2008; Bultrini, Brick et al. 2009). In particular, one of the articles presented in this thesis is largely based on the availability of this data. The intra-genome diversity discovered within the 3D7 strain has prompted the investigation of intra-population diversity: similarly to the *var* gene family, there seems to be an extensive polymorphism of *rif* and *stevor* sequences within parasite populations (Albrecht, Merino et al. 2006; Volkman, Sabeti et al. 2007; Blythe, Niang et al. 2009)

Some work has focused on the transcription of *rif* or *stevor* genes, identifying basic transcriptional elements of some *rif* genes and the antigenic switching of *rif* and *stevor* genes (Sutherland 2001; Tham, Payne et al. 2007; Blythe, Yam et al. 2008; Blythe, Niang et al. 2009; Cabral and Wunderlich 2009; Wang, Mwakalinga et al. 2010). Unfortunately, published research on these gene families is sometimes designed to correlate their results with the expression of the *var* gene family, often underestimating the importance of their basic findings regarding *rif* and *stevor* (Sharp, Lavstsen et al. 2006; Tham, Payne et al. 2007; Wang, Magistrado et al. 2008).

Our sub-division of the RIFIN family into coherent groups (Joannin, Abhiman et al. 2008) has triggered the exploration of divergent protein localizations, characteristics and transcription patterns (Petter, Haeggström et al. 2007;

Petter, Bonow et al. 2008; Bachmann, Esser et al. 2009; Bultrini, Brick et al. 2009).

With the addition of the manuscripts presented in this thesis, this summary represents the current extent of the knowledge of RIFIN and STEVOR proteins. The presence of these proteins in so many stages of the parasites life cycle as well as the high diversity within these gene families tells us that there is still much work to accomplish and many functions to discover.

1.7 PROTEIN FAMILIES, ANTIGENIC VARIANT PROTEIN FAMILIES AND PROTEIN FUNCTION(S)

Protein families: Gene duplication (Ono 1972) is a major force in the evolution of genomes, usually caused by unequal crossing over of chromosomes. A gene duplication event initially produces two identical copies of the same gene and evolutionary forces drive the fate of these genes:

- *Functional redundancy*: the two copies are maintained in order to produce large quantities of the encoded protein (Force, Lynch et al. 1999).
- *Pseudogenisation*: one of the two copies may accumulate deleterious mutations and loose its function (Wilde 1986).
- *Neofunctionalization*: one of the two copies develops a new function while the other copy maintains the original function (Prince and Pickett 2002).
- *Subfunctionalization*: the original function may be sub-divided among the duplicated genes (Force, Cresko et al. 2005).

Gene duplications, in combination with speciation events, results in complicated relationships among genes/proteins. The following definitions describe these relationships (see also Figure 3): Homology describes the common evolutionary origin of two or more traits. They are derived from a common ancestor. In genetics, the traits are protein and DNA sequences, of which sequence similarity is compared. Homologous sequences include both orthologs and paralogs. Orthology and paralogy are two terms that describe different relationships between homologous sequences (Fitch 1970).
Orthologs are sequences that derive from a single gene in the common ancestor through vertical descent and separated by speciation events.
Paralogs are homologous sequences that were separated by a gene duplication event. In this case, the two copies occupy two different positions

within the same genome.

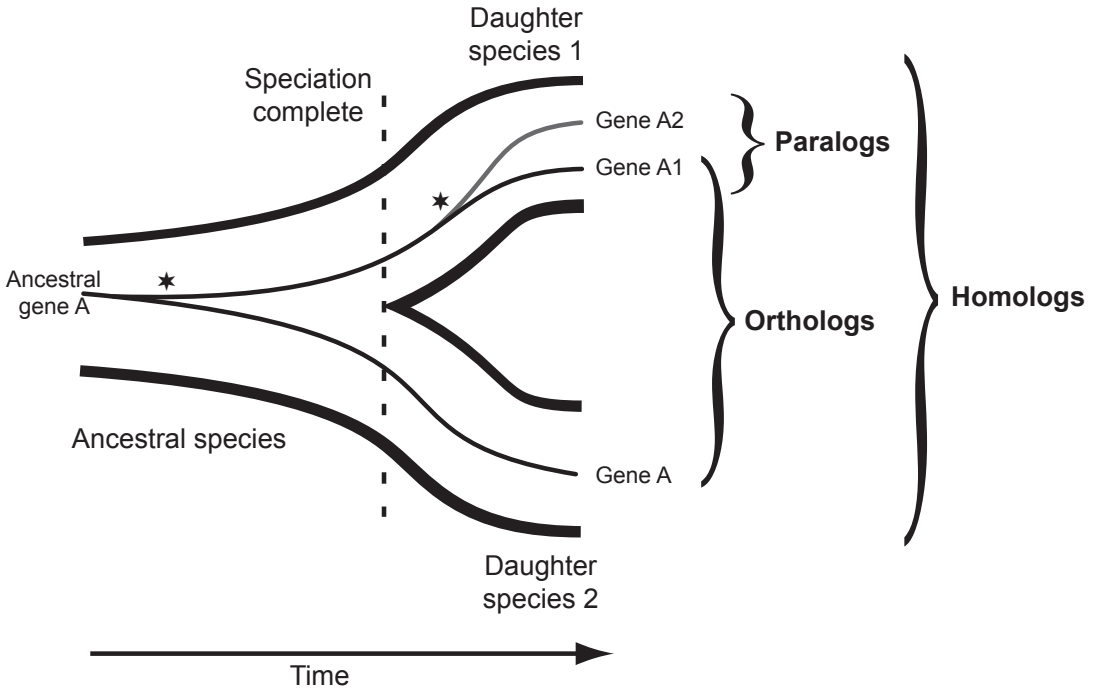

Figure 4: Relationships of duplicated genes after a speciation event. The ✶ indicates duplication events.

Distinguishing between orthologs and paralogs is important because orthologous proteins are more likely to have a similar function between two species than paralogous proteins within a species, as the latter tend to undergo sub- or neofunctionalization. However, orthology does not *imply* similarity of function. Nevertheless, homology, orthology and paralogy are commonly used in bioinformatics to assign putative functions to unknown proteins discovered through large scale sequencing projects.

Clusters of related protein sequences created by successive gene duplication and speciation events are known as a homologous protein family. Establishing a group of sequences by similarity is often a statement that they might have similar structures and in turn similar functions. Often, a protein family will have evolved to accommodate a wide range of functions, even though the general function may be the same for all the proteins (e.g. kinases with different substrate specificity). A sub-family can be viewed as a set of proteins with a related functions and domain organizations resulting from a particular line of evolution within a family.

Protein domains are fundamental units for protein evolution, typically defined as independent and globular folding units within a three-dimensional protein structure. One of the most widely used protein domain databases is Pfam (Finn, Mistry et al. 2010). In *P. falciparum*, the PfEMP1 proteins are composed of multiple domains (Lavstsen, Salanti et al. 2003), whereas the RIFIN and STEVOR proteins are only composed of one domain.

Antigenic variant protein families…: One particular fact about antigenic variant sequences is that they are usually unique to the species (or a small group of closely related species) to which they belong. For instance, *var, rif* and *stevor* genes are unique to *Plasmodium falciparum* and *P. reichenowi*. The necessary consequence is that studying an antigenic variant gene family is often synonymous with studying a paralogous gene family, as all the variants derive from gene duplication events. One should therefore be aware of possible changes in function between subsets of the protein family.

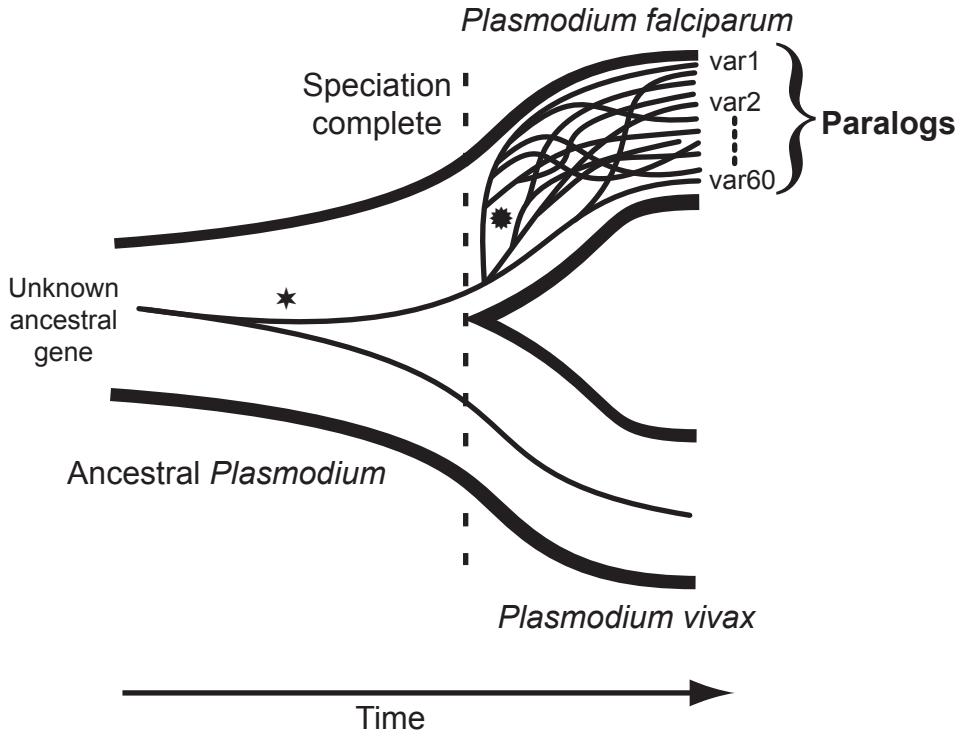

Figure 5: Expansion of a multicopy gene family in *Plasmodium*.
✹ represents multiple duplications and recombinations

The RIFIN and STEVOR proteins have been described by one Pfam domain, whereas the PfEMP1 proteins are composed of a mosaic of Duffy_binding (DBL) and CIDR domains. These groupings are useful for identifying

divergent sequences belonging to the same protein families, but they where solely based on a homology basis and not on functional basis. Different PfEMP1 variants have different binding properties to host receptors (Figure 4). The classification of these domains does not reflect for these shifts in function, which proves to be a limitation of these domain models.

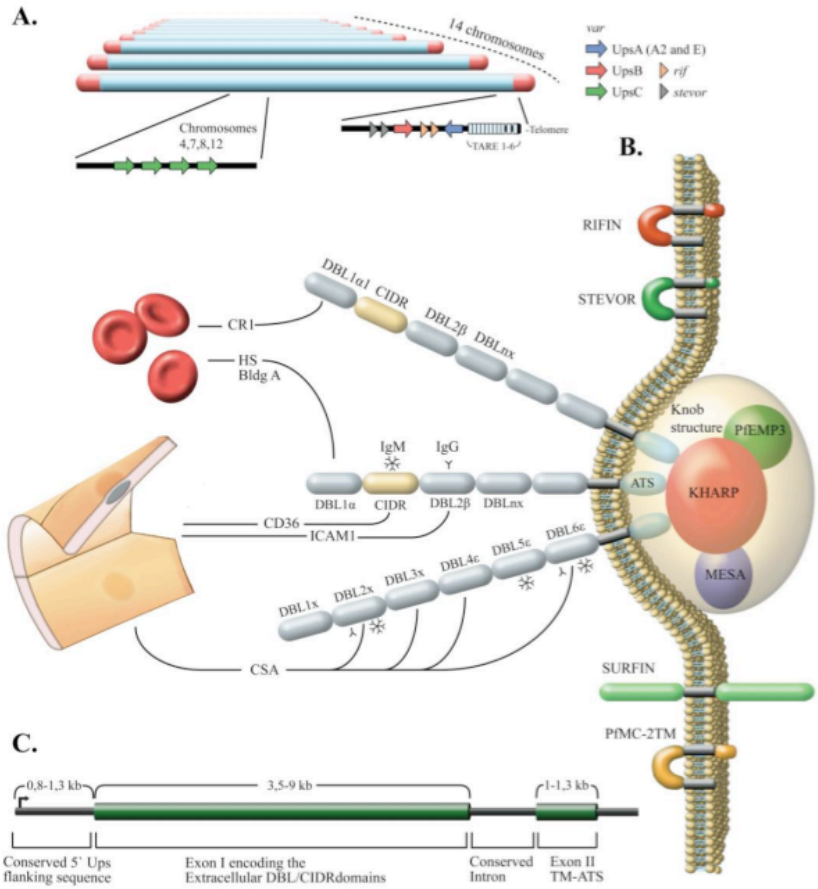

Figure 6: Organization of the PfEMP1 protein family. (A) Genomic organization of var genes in sub-telomeric and central positions. (B) PfEMP1 location on the infected RBC surface and the summary of the different binding affinities of the different domains. (C) Gene structure. Courtesy Johan Normark.

… and Protein Function: Therefore, an important step in studying antigenic variant protein families is determining whether their grouping into one family has functional relevance or not. Several studies have aimed at deciphering the antigenic variant "languages" of the DBL and CIDR domains of PfEMP1 proteins, either by looking for binding properties or by searching for associations of motifs and severity of disease (Dahlbäck, Rask et al. 2006; Rasti, Namusoke et al. 2006; Normark, Nilsson et al. 2007). These are good steps in the right direction; however, they still remain too focused on sequence diversity and motifs. Future research should aim at compiling the

empirical data obtained from individual sequences and integrate them into a model with which to compare all available data.

In the end, the main goal of grouping sequences is to infer function, not family relationships.

2 AIMS OF THE THESIS

2.1 GENERAL AIM

The general aim of this thesis is to describe and most importantly to simplify the RIFIN antigenic variant protein family.

2.2 SPECIFIC AIMS

The specific aims of the work included in this thesis are to:

i. Determine if the RIFIN protein family is a cohesive group, all members fulfilling the same function (Paper I)
ii. Develop the necessary tools to make the results of my work accessible and usable by fellow researchers (Paper II and IV)
iii. Determine the topology of the A-RIFIN and B-RIFIN proteins (Paper III)

3 RESULTS AND DISCUSSION

3.1 PAPER I: SUB-GROUPING AND SUB-FUNCTIONALIZATION OF THE RIFIN MULTI-COPY PROTEIN FAMILY

Antigenic variation is a phenomenon found in pathogenic organisms that need to avoid their host's immune system. In order to efficiently avoid the immune system, the malaria parasite *Plasmodium falciparum* has amplified several gene families, including the *var*, *stevor* and *rif* genes. The latter is the largest family of this species. However, strict mutual exclusion of protein expression has not been ascertained for this family. In effect, several publications show evidence that multiple RIFIN proteins are expressed at the same time (Fernandez, Hommel et al. 1999; Kyes, Rowe et al. 1999; Petter, Haeggström et al. 2007; Petter, Bonow et al. 2008). This apparent contradiction to the status of "antigenic variant family", as well as the large number of sequences belonging to this group, prompted us, in Paper I, to investigate this protein family.

Using phylogenetic methods, we determined that the RIFIN proteins can be classified into two major groups: A-type and B-type RIFIN proteins. Although the groups are determined by a large number of amino acids specific to each, major determinants are a 25 amino acid residue indel present in A-type and absent from B-type RIFIN proteins, as well as a different number of conserved cysteine residues, ten for A-type and five for B-type sequences. In addition to these two major groups, the B-type RIFIN proteins are further divided into three sub-groups: B1-, B2- and B3-RIFIN proteins.

Where the conserved region of B-type RIFIN is clearly distinct from the conserved region of A-type RIFIN proteins, the large variable region of B3-type sequences has similarities with that of the A-type RIFIN. We suggest this to be the result of frequent recombination events, which are expected to generate the diversity of antigenic variant proteins in *Plasmodium falciparum*, although this could also be the result of a very high mutation rate specific to the variable region of these sequences. Our analysis of 500 bp of up- and down-stream non-coding sequence also supports the sub-grouping into two major groups, although there is no congruence within the A- and B-RIFIN sub-groups.

This division of the RIFIN protein family is also substantiated by differential localizations found by others (Petter, Haeggström et al. 2007). These differences suggest that these proteins may have different functions.

To test this hypothesis, we carried out function shift analysis (Abhiman and Sonnhammer 2005) of our sub-groups. This method consists of comparing the conservation state of each amino acid between the two groups. Amino acid positions can be either variable or conserved for the whole family, in which case they do not distinguish the two groups. However, Rate Shifting Sites (RSS) and Conservation Shifting Sites (CSS) represent important determinants of function shift between sub-groups (Figure 5). The first, RSS, are amino acid position that are variable in one group but conserved in the other. The second, CSS, are positions that are conserved in both groups, but with a different amino acid. We have determined that the number of both types of shifting sites between A- and B-RIFIN proteins is consistent with a shift in their function.

Our findings allow us to simplify this large protein family into meaningful sub-groups. We expect this to facilitate the study of RIFIN proteins, in particular the discovery of their functions.

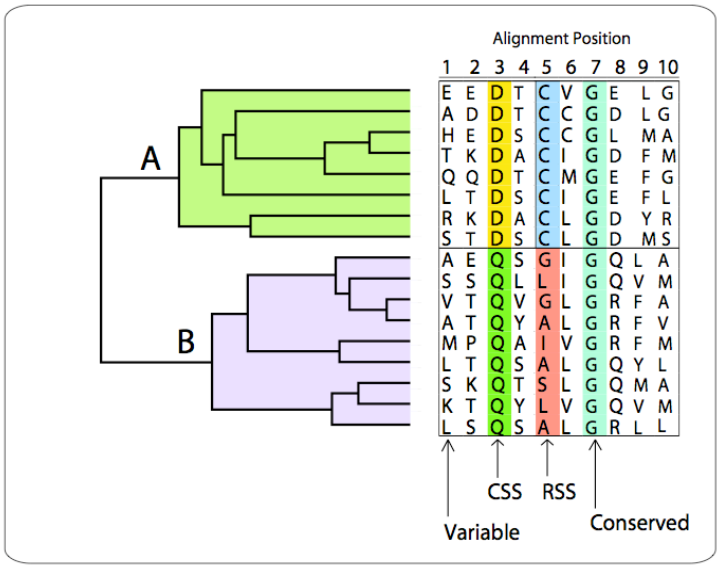

Figure 7: Schematic explanation of CSS and RSS positions. In the CSS position, the conserved amino acid of group A is aspartic acid (D) whereas in group B it is glutamine (Q). In the RSS position, the group A has a conserved cysteine (C) whereas the group B has either glycine, alanine, isoleucine or leucine (G, A, I and L, respectively). (Abhiman 2006)

3.2 PAPER II: RSPRED, A SET OF HIDDEN MARKOV MODELS TO DETECT AND CLASSIFY THE RIFIN AND STEVOR PROTEINS OF PLASMODIUM FALCIPARUM

Determining that a protein family can be divided into meaningful sub-groups is a crucial step in understanding large groups of homologous sequences. However, classifying an unknown protein needs to be accessible to a broad range of researchers for the sub-grouping to be most useful. In Paper II we describe a new tool, which we named RSpred, consisting of a set of hidden Markov models and an evaluation program to automatically detect and classify RIFIN and STEVOR proteins into their sub-groups.

To accomplish this, we needed a large number of curated RIFIN and STEVOR sequences, including sufficient proteins of each sub-group. We extracted the RIFIN and STEVOR protein sequences from the *Plasmodium falciparum* reference strain genome, 3D7, as well as those present in the Uniprot Knowledgebase. As these sequences were not sufficient, we chose to manually curate the RIFIN and STEVOR sequences from the DD2 and HB3 strains that were sequenced at the Broad Institute of Harvard and MIT. In addition to 30 new *rif* and *stevor* genes, we determined that over 25% of the automatic predictions were incorrect.

Using these newly annotated sequences as well as those from the reference strain, we revised our results from Paper I. We confirmed the sub-grouping of RIFIN proteins into A-RIFIN, B1- and B2-RIFIN sequences. However, we realized that the B3-RIFIN proteins do not have any specific characteristic that would define them as a cohesive group. We have therefore decided to rename them B-RIFIN sequences. As we had not yet investigated the STEVOR proteins, we verified the coherence of the group. The inclusion of the curated DD2 and HB3 data sets allowed us to detect a small number of divergent sequences that we designated STEVOR-like sequences.

We used our manually annotated RIFIN and STEVOR sub-groups (with the exception of the STEVOR-like) to train hidden Markov models to detect each type of sequence. A program was written to evaluate the results according to manually set cut-offs. We describe the limits of detection and the sensitivity of our method. When compared to the existing tools, RSpred proves to be more

sensitive and more specific than Pfam or TIGRFAMs, as neither sort these proteins into sub-groups.

3.3 PAPER III: RIFIN PROTEIN TOPOLOGY: A NEW PERSPECTIVE ON PLASMODIUM FALCIPARUM VARIABLE SURFACE ANTIGENS

RIFIN proteins have traditionally been described as “two transmembrane” proteins (Cheng, Cloonan et al. 1998; Kyes, Rowe et al. 1999; Lavazec, Sanyal et al. 2006). However, a recent publication from *Bultrini et al.* (Bultrini, Brick et al. 2009) challenges this statement: they describe a different protein topology between the two major RIFIN sub-groups described in Paper I. We have decided, in Paper III, to investigate A- and B-RIFIN protein topology from a bioinformatics as well as a biochemical perspective.

We applied nine different protein transmembrane prediction algorithms to two RIFIN proteins and compared their results. About half of them predict the A-RIFIN protein to have two transmembrane domains, whereas seven out of nine programs show this prediction for the B-RIFIN protein. Both proteins are predicted to have a C-terminal transmembrane region by all of the algorithms tested. However, the ambiguity lies in the middle of the protein, at the frontier between the fairly conserved and the very variable region of RIFIN proteins. This ambiguity prompted us to use a biochemical *in vitro* method to determine whether A- and B-RIFIN proteins have either one or two transmembrane regions.

We used a glycosylation mapping approach (Devoto, Piffanelli et al. 1999; van Geest, Nilsson et al. 1999; Lundin, Nordström et al. 2006) to map the location of the conserved and the variable regions of the two RIFIN proteins relative to the membrane of dog pancreas rough microsomes (i.e. endoplasmic reticulum equivalent). This method allows us to show that both the conserved and variable regions of RIFIN proteins are exposed to the lumen of the microsomes. Consequently, both A- and B-RIFIN proteins only have one transmembrane region in the C-terminus of the protein. These results also imply that RIFIN proteins expose the bulk of their sequence to the extracellular environment. Finally, we discuss the probability of different protein foldings that would allow these proteins to be exposed to the immune

system while protecting their conserved regions.

These results resolve the contradiction found in the literature. They have obvious implications on the conceptual thoughts on the function of RIFIN proteins as well as on the design of future experiments.

3.4 PAPER IV: VARDB: A PATHOGEN-SPECIFIC SEQUENCE DATABASE OF PROTEIN FAMILIES INVOLVED IN ANTIGENIC VARIATION

Antigenic variation and drift are widespread phenomena that allow pathogenic organisms to avoid their host's immune system and establish chronic infections. Although these organisms are evolutionarily diverse, they have independently developed common mechanisms to enable this ability to mutate and/or recombine their antigenic genes, and sometimes switch the expression of these genes. This results in an accumulation of genetic variability, which is a hallmark of antigenic variation. As the products of these genes are generally unique to each pathogen, research on antigenic variation is most often focused on individual species and is even, within species, usually restricted to only one of the antigenic variation gene families. This compartmentalization of research on antigenic variation often makes it difficult to find pertinent information from other gene families/pathogenic species to compare with one's own data. In addition to these difficulties, existing data from each gene family is often scattered across different databases and sometimes only included in the supplemental information provided with publications.

We developed the varDB database to serve as a repository for antigenic variant sequences, as well as their associated features. Regrouping antigenic variant data from multiple organisms in one centralized resource is also intended to make research from many multicopy gene families accessible and easily comparable.

The current varDB release (v4) includes over 83878 DNA sequences, with 46216 protein translations, 23 diseases, 32 pathogens and 46 gene families. The figure 6 briefly describes the sequence mining pipeline.

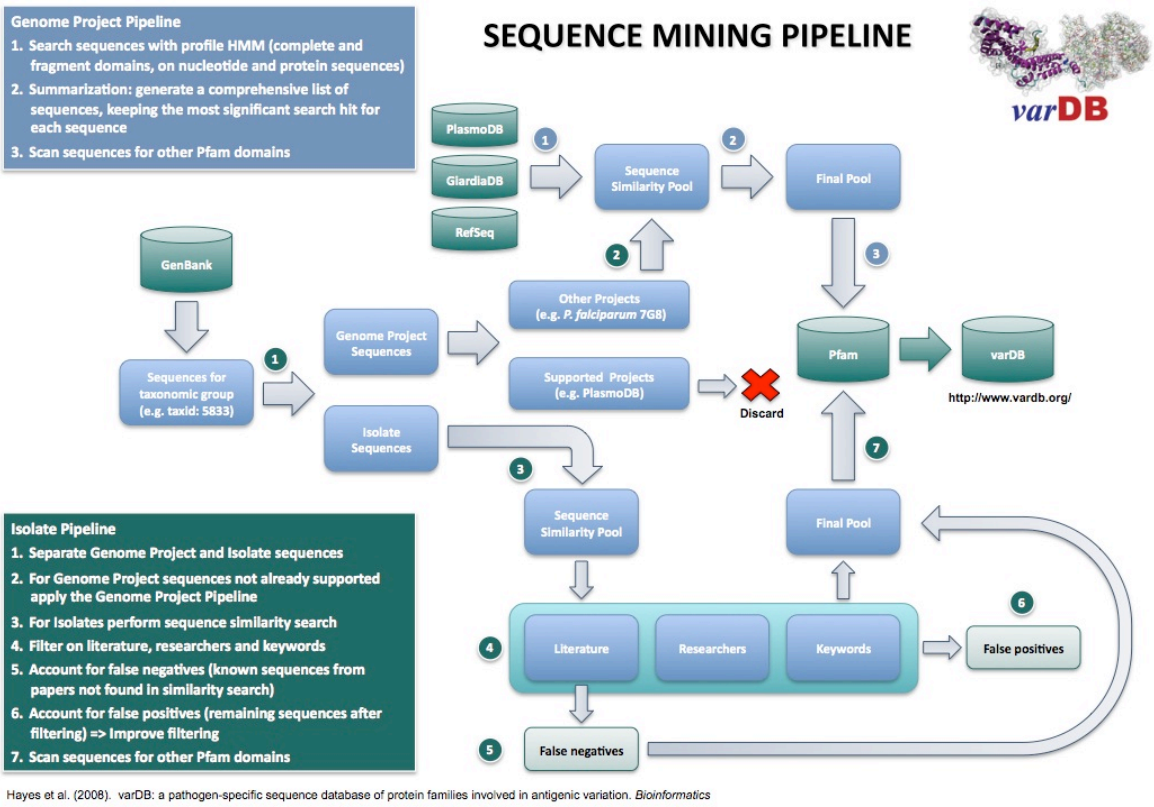

Figure 8: Sequence mining pipeline of the varDB database (Hayes, Diez et al. 2008).

4 CONCLUDING REMARKS AND PERSPECTIVES

“What is the function of RIFIN proteins?”
That is the question I was repetitively asked during the first two years of my PhD studies... Today, I am still unable to answer this question; however, through the work my collaborators and I have accomplished, I believe we have set conceptual ideas and resources that will direct future research in the right direction.

The work presented in this thesis can be summarized with the following major advancements:

- Simplification of the RIFIN family into meaningful sub-groups;
- Prediction of function shift between the two major RIFIN sub-groups;
- Development of a tool, RSpred, for the detection and classification of RIFIN and STEVOR sub-groups;
- Determination of A- and B-RIFIN protein topology;
- Development of a common resource and repository for antigenic variation sequence data.

These past two decades have seen the rise of mass sequencing, whole genome transcriptomics and proteomics and an increasing belief that bigger is better… But is it really?

The study of antigenic variant proteins families in *Plasmodium* species has unfortunately taken the route of “bigger is better, the more the merrier”. Large-scale projects result in the discovery of many previously unknown genes and proteins, some of which may be linked by homology. The result is the creation of large protein families based on the suspicion of paralogy or orthology without further proof of concept (e.g. the PIR, VIR and YIR super families (Del Portillo, Fernandez-Becerra et al. 2001; Janssen, Phillips et al. 2004; Cunningham, Jarra et al. 2005)). All of these large multicopy protein families need to be reexamined for stronger evidence of functional similarity and eventually sub-grouped accordingly.

If we want to find out how these proteins work and are involved in disease, we need to simplify them to a manageable model, where the elucidation of the properties and functions of a limited number of proteins will truly reflect the state for the larger group.

5 ACKNOWLEDGMENTS

Many people have helped me develop and grow throughout the years. I would like to thank you all here!

First and foremost, my thanks go to **Prof. Mats Wahlgren**. You provided me with an excellent working place, great colleagues and all the liberty one could ask for!

I also thank you, **Prof. Björn Andersson**, for always having an open office where I could come and voice my ideas, get some feedback and move forward!

My thanks also go to **Ass. Prof. Susumu Goto** and **Dr. Jude Przyborski** for kindly receiving me in their laboratories.

I am very grateful to all my present and past collaborators: Prof. Mats Wahlgren, Prof. Björn Andersson, Prof. Gunnar von Heijne, Prof. Bengt Persson, Prof. Erik Sonnhammer, Prof. Minoru Kanehisa, Ass. Prof. Susumu Goto, Ass. Prof. IngMarie Nilsson, Ass. Prof. Craig Wheelock, Dr. Yvonne Kallberg, Dr. Nelson Hayes, Dr. Diego Diez, Dr. Saraswathi Abhiman, Dr. Johan Normark, Oscar Franzén, Hamid Darban, Karl Enquist and Wataru Honda.

Thank you, to all the senior parasitologists: Prof. Qijun Chen, Prof. Fred Kironde, Ass. Prof. Akira Kaneko, Dr. Johan Lindh, Dr. Antonio Baragan, Dr. Susanne Nylén and Dr. Victor Fernandez.

Many thanks go to all the present members of the Wahlgren Lab: Davide Angeletti, Dr. Kirsten Moll, Karolina Palmgren, Dr. Letusa Albrecht, Dr. Kristina Persson, Sandra Nilsson, Dr. Karin Blomqvist, Mia Palmkvist, Josea Rono, Hodan Ismail, Sherwin Chan, Dr. Susanne Nylén, Dr. Srinivasulu Reddy Basi Reddy, Pilar Quintana, Jana Busch, Dr. Sriwipa, Dr. Isabel Dellacasa and Dr. Steven Kiwuwa.

In particular I would like to thank Inger, Satu and Karolina for all the help you've provided us over the years ;)

All my friends at MTC/SMI/KI:
Jessica B. (we're expecting you and Ola ;), Marlene, Linda, Sönke, Agaristi, Claudia, Emilie F., Gery, Gary, Johan A., Marianne L., Cecilia, Ewert, Lisbeth, Silvia Botero, Polya, Romanico, Ester, Jadwiga, Leigh, Sabina, Greger, Torbjörn, Tage, Marcelo, Alex, Sándor, Brigitta, Nancy, Nicolas R., Steffano, Vishal, Inga, Jakob, Eva B., Sophie M-P, Wilhelm, Isabel V., Pedro F., Lech, Rosina, Lotta, Raju and Lena…

Past lab people:
Anders B., Malin H., Fredrik P., Sanjay A., Gerd W., Anna V., Johan N., Fingani M., Ulf R., Kim B., Bobo M., Niloo R., Inger B., Satu, Johanna,

Susanna, Jon, Arnaud C., Thorsten, Irene, Anika I., Niclas…

Sabri and Suleyman, I will definitely miss coming to get lunch or diner at your gatukök!

And talking about food, Bruno & Annelie, thank you for the best galettes I've ever had! And most of all for the evenings chit-chatting about Sweden, France and sometimes politics ;)

My good friends, I will miss you all: Aurel and Vera, Kirsten and Markus M., Karolina and Markus P., Letusa and Fabio, Craig and Åsa, Gary C., Wilhelm P., Ulf M-H, Karolina E-G, Jessica and Ola, Eva B., Jane T. (have been missing you for a long time ☹), Lena L. Arnaud and Riina, Johan and Monica… I'm sure I'll see you soon ;)

Finally I would like to thank my family, and in particular, my mother Nicole. She has worked hard to give me the opportunity to study so long. I wouldn't have been able to do it without your help all those years…

And of course, Emilie, my wife… Some of you know her, others don't. I'm sure that a lot of you are surprised: "how could anyone stand that guy???" And some of you, whom I haven't seen for a while, may be surprised that I "settled down".

Let me tell you this: first, I'm not that much of a pain in the behind! Second, once you find the right person, then it is time to "settle down". And I am lucky because I found that person.

Emilie, I want to thank you for many things.
First, for letting me write my thesis without stressing (too much ;) and without bugging me (too much ;) to know how much I'd written. As you know, I'm a slow writer, but I don't need to re-write (much) once I've committed ink to the paper.
Second, for accepting to marry me ☺. As a lot of people seem to think, I will admit that I'm a "difficult" person, sometimes…
Third, for all those things I have in my heart, but that I won't write here: they belong to you and me only. You know how much I love you, and I'm sure that all the people who've known us together don't need any more proof of it.
Finally, thank you for being who you are. Je t'aime!

Disclaimer: Emilie, this list is absolutely non-exhaustive and I'll need the rest of my life to give it to you in full!

And for all those friends and colleagues I'm forgetting… Sorry, I have to run to the printer now ☹

6 REFERENCES

Abdel-Latif, M. S., K. Dietz, et al. (2003). "Antibodies to Plasmodium falciparum rifin proteins are associated with rapid parasite clearance and asymptomatic infections." Infect Immun **71**(11): 6229-6233.

Abdel-Latif, M. S., A. Khattab, et al. (2002). "Recognition of variant Rifin antigens by human antibodies induced during natural Plasmodium falciparum infections." Infect Immun **70**(12): 7013-7021.

Abhiman, S. (2006). "Prediction of function shift in protein families. ." ki.se.

Abhiman, S. and E. L. L. Sonnhammer (2005). "Large-scale prediction of function shift in protein families with a focus on enzymatic function." Proteins **60**(4): 758-768.

Albrecht, L., E. F. Merino, et al. (2006). "Extense variant gene family repertoire overlap in Western Amazon Plasmodium falciparum isolates." Mol Biochem Parasitol **150**(2): 157-165.

Bachmann, A., C. Esser, et al. (2009). "Absence of erythrocyte sequestration and lack of multicopy gene family expression in Plasmodium falciparum from a splenectomized malaria patient." PLoS ONE **4**(10): e7459.

Barry, A. E., A. Leliwa-Sytek, et al. (2007). "Population genomics of the immune evasion (var) genes of Plasmodium falciparum." PLoS Pathog **3**(3): e34.

Baruch, D. I., B. L. Pasloske, et al. (1995). "Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes." Cell **82**(1): 77-87.

Blythe, J., M. Niang, et al. (2009). "Characterization of the repertoire diversity of the Plasmodium falciparum stevor multigene family in laboratory and field isolates." Malar J **8**(1): 140.

Blythe, J. E., X. Y. Yam, et al. (2008). "Plasmodium falciparum STEVOR proteins are highly expressed in patient isolates and located in the surface membranes of infected red blood cells and the apical tips of merozoites." Infect Immun **76**(7): 3329-3336.

Boni, M. F. (2008). "Vaccination and antigenic drift in influenza." Vaccine **26 Suppl 3**: C8-14.

Bozdech, Z., M. Llinás, et al. (2003). "The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum." PLoS Biol **1**(1): E5.

Brolin, K. J. M., U. Ribacke, et al. (2009). "Simultaneous transcription of duplicated var2csa gene copies in individual Plasmodium falciparum parasites." Genome Biol **10**(10): R117.

Bull, P. C., B. S. Lowe, et al. (1999). "Antibody recognition of Plasmodium falciparum erythrocyte surface antigens in Kenya: evidence for rare and prevalent variants." Infect Immun **67**(2): 733-739.

Bull, P. C., B. S. Lowe, et al. (1998). "Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria." Nat Med **4**(3): 358-360.

Bultrini, E., K. Brick, et al. (2009). "Revisiting the Plasmodium falciparum RIFIN family: from comparative genomics to 3D-model prediction." BMC Genomics **10**: 445.

Cabral, F. and G. Wunderlich (2009). "Transcriptional memory and switching in the Plasmodium falciparum rif gene family." Mol Biochem Parasitol.

Casares, S., T.-D. Brumeanu, et al. (2010). "The RTS,S malaria vaccine." Vaccine **28**(31): 4880-4894.

Chen, Q., A. Barragan, et al. (1998). "Identification of Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1) as the rosetting ligand of the malaria parasite P. falciparum." J Exp Med **187**(1): 15-23.

Chen, Q., V. Fernandez, et al. (1998). "Developmental selection of var gene expression in Plasmodium falciparum." Nature **394**(6691): 392-395.

Cheng, Q., N. Cloonan, et al. (1998). "stevor and rif are Plasmodium falciparum multicopy gene families which potentially encode variant antigens." Mol Biochem Parasitol **97**(1-2): 161-176.
Cheng, Q., G. Lawrence, et al. (1997). "Measurement of Plasmodium falciparum growth rates in vivo: a test of malaria vaccines." Am J Trop Med Hyg **57**(4): 495-500.
Cohen, J. M., B. Moonen, et al. (2010). "How absolute is zero? An evaluation of historical and current definitions of malaria elimination." Malar J **9**: 213.
Cox-Singh, J., T. M. E. Davis, et al. (2008). "Plasmodium knowlesi malaria in humans is widely distributed and potentially life threatening." Clin Infect Dis **46**(2): 165-171.
Cox-Singh, J., J. Hiu, et al. (2010). "Severe malaria - a case of fatal Plasmodium knowlesi infection with post-mortem findings: a case report." Malar J **9**: 10.
Craig, A. and A. Scherf (2003). "Antigenic variation." 443.
Cunningham, D. A., W. Jarra, et al. (2005). "Host immunity modulates transcriptional changes in a multigene family (yir) of rodent malaria." Mol Microbiol **58**(3): 636-647.
Dahlbäck, M. A., T. S. Rask, et al. (2006). "Epitope mapping and topographic analysis of VAR2CSA DBL3X involved in P. falciparum placental sequestration." PLoS Pathog **2**(11): e124.
Daily, J. P., K. G. Le Roch, et al. (2005). "In vivo transcriptome of Plasmodium falciparum reveals overexpression of transcripts that encode surface proteins." J Infect Dis **191**(7): 1196-1203.
David, P. H., M. Hommel, et al. (1983). "Parasite sequestration in Plasmodium falciparum malaria: spleen and antibody modulation of cytoadherence of infected erythrocytes." Proc Natl Acad Sci USA **80**(16): 5075-5079.
Deitsch, K. W., S. A. Lukehart, et al. (2009). "Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens." Nat Rev Microbiol **7**(7): 493-503.
Del Portillo, H. A., C. Fernandez-Becerra, et al. (2001). "A superfamily of variant genes encoded in the subtelomeric region of Plasmodium vivax." Nature **410**(6830): 839-842.
Devoto, A., P. Piffanelli, et al. (1999). "Topology, subcellular localization, and sequence diversity of the Mlo family in plants." J Biol Chem **274**(49): 34993-35004.
Dharia, N. V., D. Plouffe, et al. (2010). "Genome-scanning of Amazonian Plasmodium falciparum shows subtelomeric instability and clindamycin-resistant parasites." Genome Res.
Fernandez, V., M. Hommel, et al. (1999). "Small, clonally variant antigens expressed on the surface of the Plasmodium falciparum-infected erythrocyte are encoded by the rif gene family and are the target of human immune responses." J Exp Med **190**(10): 1393-1404.
Ferreira, M. U., M. Zilversmit, et al. (2007). "Origins and evolution of antigenic diversity in malaria parasites." Curr Mol Med **7**(6): 588-602.
Finn, R. D., J. Mistry, et al. (2010). "The Pfam protein families database." Nucleic Acids Res **38**(Database issue): D211-222.
Fitch, W. M. (1970). "Distinguishing homologous from analogous proteins." Syst Zool **19**(2): 99-113.
Florens, L., X. Liu, et al. (2004). "Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes." Mol Biochem Parasitol **135**(1): 1-11.
Florens, L., M. P. Washburn, et al. (2002). "A proteomic view of the Plasmodium falciparum life cycle." Nature **419**(6906): 520-526.
Force, A., W. A. Cresko, et al. (2005). "The origin of subfunctions and modular gene regulation." Genetics **170**(1): 433-446.
Force, A., M. Lynch, et al. (1999). "Preservation of duplicate genes by complementary, degenerative mutations." Genetics **151**(4): 1531-1545.
Frank, M. and C. Enderes (2010). "Antigenic variation in Plasmodium falciparum: moving beyond the laboratory strains." Wien Klin Wochenschr **122 Suppl 1**: 7-10.

García, J. E., A. Puentes, et al. (2005). "Peptides from the Plasmodium falciparum STEVOR putative protein bind with high affinity to normal human red blood cells." Peptides **26**(7): 1133-1143.
Gardner, M. J., N. Hall, et al. (2002). "Genome sequence of the human malaria parasite Plasmodium falciparum." Nature **419**(6906): 498-511.
Gardner, M. J., H. Tettelin, et al. (1998). "Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum." Science **282**(5391): 1126-1132.
Goldberg, D. E. and A. F. Cowman (2010). "Moving in and renovating: exporting proteins from Plasmodium into host erythrocytes." Nat Rev Microbiol **8**(9): 617-621.
Haeggström, M., F. Kironde, et al. (2004). "Common trafficking pathway for variant antigens destined for the surface of the Plasmodium falciparum-infected erythrocyte." Mol Biochem Parasitol **133**(1): 1-14.
Haeggström, M., A. VON Euler, et al. (2007). "Characterization of Maurer's clefts in Plasmodium falciparum-infected erythrocytes." Am J Trop Med Hyg **76**(1): 27-32.
Hay, S. I., C. A. Guerra, et al. (2009). "A world malaria map: Plasmodium falciparum endemicity in 2007." PLoS Med **6**(3): e1000048.
Hayes, C., D. Diez, et al. (2008). "varDB: a pathogen-specific sequence database of protein families involved in antigenic variation." Bioinformatics.
Janssen, C. S., R. S. Phillips, et al. (2004). "Plasmodium interspersed repeats: the major multigene superfamily of malaria parasites." Nucleic Acids Res **32**(19): 5712-5720.
Joannin, N., S. Abhiman, et al. (2008). "Sub-grouping and sub-functionalization of the RIFIN multi-copy protein family." BMC Genomics **9**(1): 19.
Joergensen, L., D. C. Bengtsson, et al. (2010). "Surface Co-Expression of Two Different PfEMP1 Antigens on Single Plasmodium falciparum-Infected Erythrocytes Facilitates Binding to ICAM1 and PECAM1." PLoS Pathog **6**(9).
Kaviratne, M., S. Khan, et al. (2002). "Small variant STEVOR antigen is uniquely located within Maurer's clefts in Plasmodium falciparum-infected red blood cells." Eukaryotic Cell **1**(6): 926-935.
Khattab, A., I. Bonow, et al. (2008). "Plasmodium falciparum variant STEVOR antigens are expressed in merozoites and possibly associated with erythrocyte invasion." Malar J **7**: 137.
Krief, S., A. A. Escalante, et al. (2010). "On the diversity of malaria parasites in African apes and the origin of Plasmodium falciparum from Bonobos." PLoS Pathog **6**(2): e1000765.
Kyes, S. A., J. A. Rowe, et al. (1999). "Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with Plasmodium falciparum." Proc Natl Acad Sci USA **96**(16): 9333-9338.
Lambert, H. and A. Barragan (2010). "Modelling parasite dissemination: host cell subversion and immune evasion by Toxoplasma gondii." Cell Microbiol **12**(3): 292-300.
Lavazec, C., S. Sanyal, et al. (2006). "Hypervariability within the Rifin, Stevor and Pfmc-2TM superfamilies in Plasmodium falciparum." Nucleic Acids Res **34**(22): 6696-6707.
Lavazec, C., S. Sanyal, et al. (2007). "Expression switching in the stevor and Pfmc-2TM superfamilies in Plasmodium falciparum." Mol Microbiol **64**(6): 1621-1634.
Lavstsen, T., A. Salanti, et al. (2003). "Sub-grouping of Plasmodium falciparum 3D7 var genes based on sequence analysis of coding and non-coding regions." Malar J **2**: 27.
Le Roch, K. G., J. R. Johnson, et al. (2004). "Global analysis of transcript and protein levels across the Plasmodium falciparum life cycle." Genome Res **14**(11): 2308-2318.
Le Roch, K. G., Y. Zhou, et al. (2002). "Monitoring the chromosome 2 intraerythrocytic transcriptome of Plasmodium falciparum using oligonucleotide arrays." Am J Trop Med Hyg **67**(3): 233-243.
Le Roch, K. G., Y. Zhou, et al. (2003). "Discovery of gene function by expression profiling of the malaria parasite life cycle." Science **301**(5639): 1503-1508.

Limpaiboon, T., D. W. Taylor, et al. (1990). "Characterization of a Plasmodium falciparum epitope recognized by a monoclonal antibody with broad isolate and species specificity." Southeast Asian J Trop Med Public Health **21**(3): 388-396.
Llinás, M., Z. Bozdech, et al. (2006). "Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains." Nucleic Acids Res **34**(4): 1166-1173.
Lundin, C., R. Nordström, et al. (2006). "Membrane topology of the human seipin protein." FEBS Lett **580**(9): 2281-2284.
MacPherson, G. G., M. J. Warrell, et al. (1985). "Human cerebral malaria. A quantitative ultrastructural analysis of parasitized erythrocyte sequestration." Am J Pathol **119**(3): 385-401.
McRobert, L., P. Preiser, et al. (2004). "Distinct trafficking and localization of STEVOR proteins in three stages of the Plasmodium falciparum life cycle." Infect Immun **72**(11): 6597-6602.
Morrison, L. J., L. Marcello, et al. (2009). "Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity." Cell Microbiol **11**(12): 1724-1734.
Mumford, J. A. (2007). "Vaccines and viral antigenic diversity." Rev - Off Int Epizoot **26**(1): 69-90.
Niang, M., X. Yan Yam, et al. (2009). "The Plasmodium falciparum STEVOR Multigene Family Mediates Antigenic Variation of the Infected Erythrocyte." PLoS Pathog **5**(2): e1000307.
Normark, J., D. Nilsson, et al. (2007). "PfEMP1-DBL1{alpha} amino acid motifs in severe disease states of Plasmodium falciparum malaria." Proc Natl Acad Sci USA **104**(40): 15835-15840.
Ono, S. (1972). "Gene duplication, mutation load, and mammalian genetic regulatory systems." J Med Genet **9**(3): 254-263.
Petter, M., I. Bonow, et al. (2008). "Diverse Expression Patterns of Subgroups of the rif Multigene Family during Plasmodium falciparum Gametocytogenesis." PLoS ONE **3**(11): e3779.
Petter, M., M. Haeggström, et al. (2007). "Variant proteins of the Plasmodium falciparum RIFIN family show distinct subcellular localization and developmental expression patterns." Mol Biochem Parasitol **156**(1): 51-61.
Pongponratn, E., M. Riganti, et al. (1985). "Electron microscopy of the human brain in cerebral malaria." Southeast Asian J Trop Med Public Health **16**(2): 219-227.
Prince, V. E. and F. B. Pickett (2002). "Splitting pairs: the diverging fates of duplicated genes." Nat Rev Genet **3**(11): 827-837.
Przyborski, J. M., S. K. Miller, et al. (2005). "Trafficking of STEVOR to the Maurer's clefts in Plasmodium falciparum-infected erythrocytes." EMBO J **24**(13): 2306-2317.
Rasti, N., F. Namusoke, et al. (2006). "Nonimmune immunoglobulin binding and multiple adhesion characterize Plasmodium falciparum-infected erythrocytes of placental origin." Proc Natl Acad Sci USA **103**(37): 13795-13800.
Rottmann, M., C. McNamara, et al. (2010). "Spiroindolones, a potent compound class for the treatment of malaria." Science **329**(5996): 1175-1180.
Rowe, J. A., J. M. Moulds, et al. (1997). "P. falciparum rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1." Nature **388**(6639): 292-295.
Sachs, J. and P. Malaney (2002). "The economic and social burden of malaria." Nature **415**(6872): 680-685.
Sachs, J. D. (2002). "A new global effort to control malaria." Science **298**(5591): 122-124.
Sam-Yellowe, T. Y., L. Florens, et al. (2004). "A Plasmodium gene family encoding Maurer's cleft membrane proteins: structural properties and expression profiling." Genome Res **14**(6): 1052-1059.
Sander, A. F., A. Salanti, et al. (2009). "Multiple var2csa-type PfEMP1 genes located at different chromosomal loci occur in many Plasmodium falciparum isolates." PLoS ONE **4**(8): e6667.
Scherf, A., R. Hernandez-Rivas, et al. (1998). "Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during

intra-erythrocytic development in Plasmodium falciparum." EMBO J **17**(18): 5418-5426.
Schreiber, N., N. Brattig, et al. (2006). "Cerebral malaria is associated with IgG2 and IgG4 antibody responses to recombinant Plasmodium falciparum RIFIN antigen." Microbes Infect **8**(5): 1269-1276.
Schreiber, N., A. Khattab, et al. (2008). "Expression of Plasmodium falciparum 3D7 STEVOR proteins for evaluation of antibody responses following malaria infections in naive infants." Parasitology **135**(2): 155-167.
Sharp, S., T. Lavstsen, et al. (2006). "Programmed transcription of the var gene family, but not of stevor, in Plasmodium falciparum gametocytes." Eukaryotic Cell **5**(8): 1206-1214.
Sherman, I. W., S. Eda, et al. (2003). "Cytoadherence and sequestration in Plasmodium falciparum: defining the ties that bind." Microbes Infect **5**(10): 897-909.
Silamut, K., N. H. Phu, et al. (1999). "A quantitative analysis of the microvascular sequestration of malaria parasites in the human brain." Am J Pathol **155**(2): 395-410.
Simon, M. C. and H. J. Schmidt (2007). "Antigenic variation in ciliates: antigen structure, function, expression." J Eukaryot Microbiol **54**(1): 1-7.
Sinden, R. E. (2010). "A biologist's perspective on malaria vaccine development." Hum Vaccin **6**(1): 3-11.
Smith, J. D., C. E. Chitnis, et al. (1995). "Switches in expression of Plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes." Cell **82**(1): 101-110.
Snow, R. W. and K. Marsh (2010). "Malaria in Africa: progress and prospects in the decade since the Abuja Declaration." Lancet **376**(9735): 137-139.
Su, X. Z., V. M. Heatwole, et al. (1995). "The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes." Cell **82**(1): 89-100.
Sutherland, C. J. (2001). "Stevor transcripts from Plasmodium falciparum gametocytes encode truncated polypeptides." Mol Biochem Parasitol **113**(2): 331-335.
Tanner, M. and D. de Savigny (2008). "Malaria eradication back on the table." Bulletin of the World Health Organization (BLT) **86**(2): 81-160.
Tham, W.-H., P. D. Payne, et al. (2007). "Identification of basic transcriptional elements required for rif gene transcription." Int J Parasitol **37**(6): 605-615.
Tsarukyanova, I., J. Drazba, et al. (2009). "Proteins of the Plasmodium falciparum two transmembrane Maurer's cleft protein family, PfMC-2TM, and the 130 kDa Maurer's cleft protein define different domains of the infected erythrocyte intramembranous network." Parasitol Res **104**(4): 875-891.
Turner, C. M. R. (2002). "A perspective on clonal phenotypic (antigenic) variation in protozoan parasites." Parasitology **125 Suppl**: S17-23.
Turner, G. D., H. Morrison, et al. (1994). "An immunohistochemical study of the pathology of fatal malaria. Evidence for widespread endothelial activation and a potential role for intercellular adhesion molecule-1 in cerebral sequestration." Am J Pathol **145**(5): 1057-1069.
Udeinya, I. J., L. H. Miller, et al. (1983). "Plasmodium falciparum strain-specific antibody blocks binding of infected erythrocytes to amelanotic melanoma cells." Nature **303**(5916): 429-431.
van der Woude, M. W. and A. J. Bäumler (2004). "Phase and antigenic variation in bacteria." Clin Microbiol Rev **17**(3): 581-611, table of contents.
van Geest, M., I. Nilsson, et al. (1999). "Insertion of a bacterial secondary transport protein in the endoplasmic reticulum membrane." J Biol Chem **274**(5): 2816-2823.
Volkman, S. K., P. C. Sabeti, et al. (2007). "A genome-wide map of diversity in Plasmodium falciparum." Nat Genet **39**(1): 113-119.
Wang, C., P. Magistrado, et al. (2008). "Preferential transcription of conserved rif genes in two phenotypically distinct Plasmodium falciparum parasite lines." Int J Parasitol.
Wang, C. W., S. B. Mwakalinga, et al. (2010). "Identification of a major rif transcript common to gametocytes and sporozoites of Plasmodium falciparum." Malar J **9**(1): 147.

Weber, J. L. (1988). "Interspersed repetitive DNA from Plasmodium falciparum." Mol Biochem Parasitol **29**(2-3): 117-124.
WHO (2008). "World Malaria Report 2008." WHO Press: 190.
Wilde, C. D. (1986). "Pseudogenes." CRC Crit Rev Biochem **19**(4): 323-352.
Wisniewski-Dyé, F. and L. Vial (2008). "Phase and antigenic variation mediated by genome modifications." Antonie Van Leeuwenhoek.